UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
10/647,203	08/21/2003	Alexander Franz	24207-10274	1475

62296   7590   08/24/2007

GOOGLE / FENWICK
SILICON VALLEY CENTER
801 CALIFORNIA ST.
MOUNTAIN VIEW, CA 94041

EXAMINER
SHAH, PARAS D

ART UNIT	PAPER NUMBER
2626	

MAIL DATE	DELIVERY MODE
08/24/2007	PAPER

Please find below and/or attached an Office communication concerning this application or proceeding.

The time period for reply, if any, is set in the attached communication.

-- *The MAILING DATE of this communication appears on the cover sheet with the correspondence address --*

Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE <u>3</u> MONTH(S) OR THIRTY (30) DAYS, WHICHEVER IS LONGER, FROM THE MAILING DATE OF THIS COMMUNICATION.
- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

1)☒ Responsive to communication(s) filed on <u>*10 July 2007*</u>.

2a)☒ This action is **FINAL.**   2b)☐ This action is non-final.

3)☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims

4)☒ Claim(s) *1, 3-6, 8-14, 16-25, 27-36* is/are pending in the application.

    4a) Of the above claim(s) _____ is/are withdrawn from consideration.

5)☐ Claim(s) _____ is/are allowed.

6)☒ Claim(s) *1,3,4,8,9,11-14,16-18,20-24,27-29 and 31-36* is/are rejected.

7)☒ Claim(s) *5,10,19,25 and 30* is/are objected to.

8)☐ Claim(s) _____ are subject to restriction and/or election requirement.

Application Papers

9)☐ The specification is objected to by the Examiner.

10)☐ The drawing(s) filed on _____ is/are: a)☐ accepted or b)☐ objected to by the Examiner.

    Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).

    Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).

11)☐ The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

Priority under 35 U.S.C. § 119

12)☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).

    a)☐ All b)☐ Some * c)☐ None of:

        1.☐ Certified copies of the priority documents have been received.

        2.☐ Certified copies of the priority documents have been received in Application No. _____.

        3.☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

    * See the attached detailed Office action for a list of the certified copies not received.

Attachment(s)

1) ☒ Notice of References Cited (PTO-892)
2) ☐ Notice of Draftsperson's Patent Drawing Review (PTO-948)
3) ☐ Information Disclosure Statement(s) (PTO/SB/08)
    Paper No(s)/Mail Date _____.

4) ☐ Interview Summary (PTO-413)
    Paper No(s)/Mail Date. _____.
5) ☐ Notice of Informal Patent Application
6) ☐ Other: _____.

DETAILED ACTION

1.   This communication is in response to the Amendments and Arguments filed on

07/10/2007. Claims 1, 3-6, 8-14, 16-25, 27-36 remain pending and have been

examined. The Applicants' amendment and remarks have been carefully considered,

but they are not persuasive and do not place the claims in condition for allowance.

Accordingly, this action has been made FINAL.

2.   All previous objections and rejections directed to the Applicant's disclosure and

claims not discussed in this Office Action have been withdrawn by the Examiner.

Change of Art Units

3.   It should be note that the Examiner has changed art units, which was formerly

2112. The Examiner's new art unit is 2626.

Response to Arguments

4.   Applicant's arguments (pages 12-18) filed on 07/10/2007 with regard to 1, 3-6, 8-

14, 16-25, 27-36 have been fully considered but they are not persuasive and are moot

in view of new grounds for rejection. Due to the newly added limitations, a new

reference was applied. The added limitations comprise the language "configured to" and

"subset" as seen in claims 1 and 13. Further, the limitation "at least one token" in claim

6 and the limitation "at least in part provided in a vocabulary" as recited in claim 24.

        Further, it should be noted that the Applicant's argument that Su does not

contain a vocabulary and the cited portion shows distribution statistics is traversed by

the examiner. It was pointed out that in Su *et al.* page 24, 2nd full paragraph, sect.

Simulation, (1st paragraph), line 5-8 that there is a compound list that is updated based

on the likelihood value. Hence, the mentioned reference implicitly teaches the

vocabulary being present.

Response to Amendment

6.   Applicants' amendments filed on 07/10/2007have been fully considered. The

newly amended limitations in claims 1, 3-6, 8-14, 16-25, 27-36 necessitates new

grounds of rejection.

Claim Rejections - 35 USC § 112

7.   The following is a quotation of the second paragraph of 35 U.S.C. 112:

> The specification shall conclude with one or more claims particularly pointing out and distinctly
> claiming the subject matter which the applicant regards as his invention.

8.   Claims 1, 3-5, 13, 15-23 are rejected under 35 U.S.C. 112, second paragraph, as

being indefinite for failing to particularly point out and distinctly claim the subject matter

which applicant regards as the invention.

9.   As to claims 1 and 13, the limitation "configured" is held to be indefinite since it

suggests optional language. See MPEP 2111.04.

10. Claims 3-5 and 15-23 are rejected as being based upon an indefinite base claim.

Claim Rejections - 35 USC § 103

10.   The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all

obviousness rejections set forth in this Office action:

> (a) A patent may not be obtained though the invention is not identically teach or described as set forth
> in section 102 of this title, if the differences between the subject matter sought to be patented and the
> prior art are such that the subject matter as a whole would have been obvious at the time the invention
> was made to a person having ordinary skill in the art to which said subject matter pertains.
> Patentability shall not be negatived by the manner in which the invention was made.

11.   Claims 1, 3, 6, 8, 11-13, 20-24, and 31-36 are rejected under 35 U.S.C. 103(a)

as being unpatentable over Su *et al.* (In *Proceedings of the 32nd Annual Meeting on*

Association For Computational Linguistics 1994) in view of Jurafsky *et al.* (Speech and

Language Processing: An Introduction to Natural Language Processing, Computational

Linguistics, and Speech Recognition).

As to claims 1, 6, and 12, Su *et al.* teaches a system for finding compound words

in a text corpus comprising:

a vocabulary (see page 24, 2nd full paragraph, sect. Simulation, (1st

paragraph), line 5-8) comprising of tokens (see page 244, Table 1) from a text

corpus (see page 243, left column, 2nd paragraph, line 6)

compound finder iteratively finding compounds having a plurality of length

within the text corpus, each compound comprising a plurality of tokens,

comprising: (page 244, left column, 1st paragraph, line 10) (e.g. It should be

noted that windowing the corpus in sizes of 2 and 3 over the text corpus can be

interpreted as a form of iteration when finding compounds of these various

lengths)

n-gram counter (see page 244, left column, 1st paragraph, lines 3-4) evaluating a frequency of occurrence (n-gram counter) (see page 244, left column, 1st paragraph, lines 3-4) for one or more n-grams (see page 243, left column, 3rd paragraph, lines 1-5) and

a likelihood evaluator to determine a likelihood of collocation for one or more of the n-grams having the same length compounds (see page 243, right column, line 8), adding a subset of n-grams having a high likelihood as compounds to the vocabulary and rebuilding the vocabulary based on the added, which adds the compound words having a high likelihood to the vocabulary (see page 245, right column, 2nd paragraph, line 7) (e.g. A subset can be 0 or more and this is done by the reference.)

Su *et al.* does not specifically teach the use of a iterator for selecting n-grams having a length that is less than the selected n-gram.

However, Jurafsky *et al.* does teach the use of an iterator configured to select n-grams having a same length that is less than a length of n-grams selected during a previous iteration (see page 216, sect. 6.4, equation 6.30, and 2nd paragraph) (e.g. From the cited reference it is seen that the n-gram starts from some maximum limit and then proceeds to a lower order n-gram no frequency count is obtained. Hence, it iterates one less than the previous length based on the current length.)

It would have been obvious to one of ordinary skilled in the art at the time the invention was made to have modified the finding of compounds words in a

corpus as taught by Su *et al.* with the backward iteration as taught by Jurafsky *et al.* The motivation to have combined the references involves the ability to solve the problem when zero frequency n-grams (see Jurafsky, page 216, sect. 6.4, equation 6.30, and 2nd paragraph). are obtained as would benefit the compound word finding as taught by Su *et al.*

As to claims 3 and 8, Su *et al.* in view of Jurafsky *et al.* teach all of the limitations as claim 1 above.

Furthermore Su *et al.* teaches a system where only some of the subset of n-grams that have a high likelihood are added as compounds to the vocabulary (page 245, right column, 2nd paragraph, line 6-8) (e.g. It should be noted that the selection of those compounds, which have a high likelihood will be chosen if the value is greater than 0, otherwise it will not be included).

As to claim 11, , Su *et al.* in view of Jurafsky *et al.* teach all of the limitations as in claim 6 above.

However, Su *et al.* in view of Jurafsky *et al.* do not specifically teach the use of a computer for compound extraction. Su *et al.* does mention simulation for compound extraction (see Su *et al.* page 245, right column, 2nd paragraph). Hence, it is obvious to one of ordinary skilled in the art to have used a computer to execute the simulation from code. The motivation to include a computer-

storage medium is for use in machine translation (see Su *et al.* page 243, left column, 1st paragraph, line 27).

As to claims 13, 24, and 36 Su *et al.* teaches a system for identifying compounds through iterative analysis comprising:

a compound finder evaluating compounds in a text corpus comprising:

n-gram counter (see Su *et al.* page 244, left column, 1st paragraph, lines 3-4) for determining number of occurrences of one or more n-grams (e.g. The maximum number of tokens depends on the iteration value or step) the number of tokens up to the limit for iteration (see Su *et al.* page 243, left column, 2nd paragraph, line 3 and line 10) (e.g. A limit is pre-specified by the reference), which are at least in part provided in a vocabulary for the text corpus (see page 244, Table 1) from a text corpus(see page 24, 2nd full paragraph, sect. Simulation, (1st paragraph), line 5-8)

a likelihood evaluator (see Su *et al.* page 243, right column, line 8), which determines a measure of association between tokens (see Su *et al.* page 243, right column, lines 20-23) and , which adds the compound words having a high likelihood to the vocabulary (see Su *et al.* page 245, right column, 2nd paragraph, line 7). Further, the adjustment of the limit can also be interpreted as the change in the n value of an n-gram. Thus, the change of limit from n=2 to n=3, will change the number of tokens per compound (see Su *et al.* page 243, left column, 2nd paragraph, lines 9-10). However, Su *et al.* does not specifically teach the use

of a stored limit of the number of tokens per compound and the use of a

vocabulary. It would have been obvious to one of ordinary skilled in the art to

have included a predetermined limit on the number of token per compound. The

motivation to modify the compound extraction by Su *et al.* by the inclusion of a

stored limit is to acquire the compounds of interest to the user (see Su *et al.*

page 243, 2nd paragraph, line 6) (e.g. The reference uses n-grams of n=2, and

n=3).

Su *et al.* does not specifically teach the use of a iterator for selecting n-

grams having a length that is less than the selected n-gram.

However, Jurafsky et al. does teach the use of an iterator configured to

select n-grams having a same length that is less than a length of n-grams

selected during a previous iteration (see page 216, sect. 6.4, equation 6.30, and

2nd paragraph) (e.g. From the cited reference it is seen that the n-gram starts

from some maximum limit and then proceeds to a lower order n-gram no

frequency count is obtained. Hence, it iterates one less than the previous length

based on the current length.)

It would have been obvious to one of ordinary skilled in the art at the time

the invention was made to have modified the finding of compounds words in a

corpus as taught by Su *et al.* with the backward iteration as taught by Jurafsky *et*

al. The motivation to have combined the references involves the ability to solve

the problem when zero frequency n-grams (see Jurafsky, page 216, sect. 6.4,

equation 6.30, and 2nd paragraph). are obtained as would benefit the compound word finding as taught by Su et al.

As to claims 20-21 and 31-32, Su et al. and Jurafsky et al. teach all of the limitations as claim 13 above.

Furthermore Su et al. teaches an initial vocabulary (see page 24, 2nd full paragraph, sect. Simulation, (1st paragraph), line 5-8) where token are extracted from a text corpus (see page 243, left column, 2nd paragraph, lines 6-9) through morphological analysis (e.g. It should be noted that morphological analysis and parsing is similar).

As to claims 22 and 33, Su et al. and Jurafsky et al. teach all of the limitations as claim 13 above.

Furthermore Su et al. teaches a filter determining the number of occurrences of one or more n-grams within the text corpus for unique n-grams (see page 243, left column, 1st paragraph, line 3 and lines 7-9) (e.g. It should be noted that the use of the relative frequency is a measure for compound extraction and can thus be interpreted as a filtering means when the compound filtering is done) (see page 243, left column, 1st paragraph, lines1-5).

As to claims 23 and 34, Su et al. and Jurafsky et al. teach all of the limitations as claim 13 above.

Furthermore Su *et al.* teaches a system where the text corpus comprises of documents comprising one of a news message and text (see Su *et al.* abstract).

As to claim 35, Su *et al.* in view of Jurafsky *et al.* teach do not specifically teach the use of a computer for compound extraction. Su *et al.* does mention simulation for compound extraction (see Su *et al.* page 245, right column, 2nd paragraph). Hence, it is obvious to one of ordinary skilled in the art to have used a computer to execute the simulation from code. The motivation to include a computer-storage medium is for use in machine translation (see Su *et al.* page 243, left column, 1st paragraph, line 27).

12.   Claims 4, 9, 16-18 and 27-29 is rejected under 35 U.S.C. 103(a) as being unpatentable over Su *et al.* in view of Jurafsky *et al.* as applied to claims 1, 6, 13, and 24 above, and further in view of Manning (The MIT Press 1999).

As to claims 4, 9, 16-17, 27, and 28 Su *et al.* in view of Jurafsky *et al.* teaches all of the limitations as in claims 1, 6, 13, and 24 above.

Furthermore, Su *et al.*, teaches where the likelihood ratio λ is computed by: $\lambda=(P(x\_|M_c)*P(M_c))/(P(x\_|M_{nc})*P(M_{nc}))$(see Su *et al.*, page 243, right column, line 9 (equation)) (e.g. It should be noted that the reference uses a different notation, but the same result and definitions are used, where the numerator is the n-gram produced by a compound result and the denominator is the result

produced by a non-compound result. The formula can be changed to account for various distributions (Gaussian, Binomial).

However, Su *et al.* in view of Jurafsky *et al.* do not specifically teach the likelihood ratio given by $\lambda = L(H_i)/L(H_c)$.

Manning shows the use of the likelihood ratio (see equation 5.10)(e.g. The equation in given in log form. The logs can be omitted to obtain the desired formula. The numerator is the independent hypothesis and the denominator is the dependence hypothesis.)

It would have been obvious to one of ordinary skilled in the art to have modified finding of compounds in a text corpus as taught by Su *et al.* and Jurafsky *et al.* with the formula as taught by Manning. The motivation to modify the former is for collocation discovery (see Manning, page 172, sect. 5.3.4, 3rd paragraph, lines 1-4).

As to claims 18 and 29, Su *et al.* in view of Jurafsky *et al.* teaches all of the limitations as claim 13, 16, and 17 above.

Su *et al.* in view of Jurafsky teach a system for identifying compounds through measure of association.

However, Su *et al.* in view of Jurafsky do not specifically teach the representation of the independence and collocation hypothesis.

Manning does teach the explanations of these two types of hypothesis (see page 172, sect. 5.3.4, bullet items) (e.g. It should be noted that the

independence hypothesis is given by hypothesis 1 and the dependence or

collocation hypothesis by hypothesis 2. The w_2 and w_1 can be interpreted as the

tokens since the reference deals with a text corpus).

It would have been obvious to one of ordinary skilled in the art to have

modified the finding of compound words in a text corpus as taught by Su *et al.*

and Jurafsky *et* al with the inclusion of the two hypothesis as taught by Manning.

The motivation to modify the former is for collocation discovery (see Manning,

page 172, sect. 5.3.4, 3rd paragraph, lines 1-4). Further, the use of the formula

presented by Manning would require an explanation of frequency for each type of

hypothesis in order to find the likelihood ratio (definition of likelihood ratio).

Allowable Subject Matter

13.   Claims 5,10, 19, 25, and 30 are objected to as being dependent upon a rejected

base claim, but would be allowable if rewritten in independent form including all of the

limitations of the base claim and any intervening claims.

14.   Claim 14 would be allowable if rewritten to overcome the rejection(s) under 35

U.S.C. 112, 2nd paragraph, set forth in this Office action and to include all of the

limitations of the base claim and any intervening claims.

15.   The following is a statement of reasons for the indication of allowable subject

matter: none of the prior art references alone or in combination teaches or fairly

suggests the limitations where "a limiter identifying a number of n-grams up to the upper

limit based on number of occurrences" as seen in claims 14 and 25. Also, the limitations

of "dividing the *n*-gram into *n*-1 pairings of segments... selecting the maximum

likelihood of collocation of the pairings as $L(H_c)$" as seen in claims 5 and 10. Further, the

limitations "$L(H_i)$ is computed ... in accordance with the formula:

$$\underset{L(H_i)}{\arg\max} \frac{L(t_1, t_2 \, formcompound)}{L(n-gramdoesnotformcompound)}$$ " as seen in claims 19 and 30.

Conclusion

16.   Applicant's amendment necessitated the new ground(s) of rejection presented in

this Office action. Accordingly, **THIS ACTION IS MADE FINAL**. See MPEP

§ 706.07(a). Applicant is reminded of the extension of time policy as set forth in 37

CFR 1.136(a).

A shortened statutory period for reply to this final action is set to expire THREE

MONTHS from the mailing date of this action. In the event a first reply is filed within

TWO MONTHS of the mailing date of this final action and the advisory action is not

mailed until after the end of the THREE-MONTH shortened statutory period, then the

shortened statutory period will expire on the date the advisory action is mailed, and any

extension fee pursuant to 37 CFR 1.136(a) will be calculated from the mailing date of

the advisory action. In no event, however, will the statutory period for reply expire later

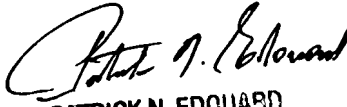than SIX MONTHS from the date of this final action.

Any inquiry concerning this communication or earlier communications from the

examiner should be directed to Paras Shah whose telephone number is (571)270-1650.

The examiner can normally be reached on MON.-THURS. 7:30a.m.-4:00p.m. EST.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Patrick Edouard can be reached on (571)272-7603. The fax phone number for the organization where this application or proceeding is assigned is 571-273-8300.

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see http://pair-direct.uspto.gov. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free). If you would like assistance from a USPTO Customer Service Representative or access to the automated information system, call 800-786-9199 (IN USA OR CANADA) or 571-272-1000.

P.S.

08/16/2007

PATRICK N. EDOUARD
SUPERVISORY PATENT EXAMINER